

AWARD NUMBER: W81XWH-12-1-0298

TITLE: MTHFR Functional Polymorphism C677T and Genomic Instability in the Etiology of Idiopathic Autism in Simplex Families

PRINCIPAL INVESTIGATOR: Xudong Liu, PhD

CONTRACTING ORGANIZATION: Queen's University
Kingston Ontario K7M 8A6 Canada

REPORT DATE: December 2014

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE December 2014		2. REPORT TYPE Final		3. DATES COVERED 30 Sep 2012 - 29 Sep 2014	
4. TITLE AND SUBTITLE MTHFR Functional Polymorphism C677T and Genomic Instability in the Etiology of Idiopathic Autism in Simplex Families				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-12-1-0298	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Xudong Liu, PhD Email: liux@queensu.ca				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Queen's University Kingston Ontario K7M 8A6 Canada				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Autism Spectrum Disorder (ASD) are a group of neurodevelopmental disorders that are caused by a range of factors, including: genetic, epigenetic and environmental, with a genetic/epigenetic model proposed (Jiang et al., 2004). While a main focus of autism research remains on the genetic causes, more and more attention was drawn to the role epigenetic factors play, as it has been shown to play a role in idiopathic autism. With our previous published study revealed significantly association of C677T polymorphism in MTHFR gene with idiopathic autism in Simplex (SPX) autism families (Liu et al., 2011); and the proven facts that de novo CNVs rates are consistently high in SPX ASD (5.8%-10.2%) versus familial ASD (2-3%), we hypothesize that low-activity MTHFR 677T allele leads to increase global DNA hypomethylation and consequently results in increased generation of de novo CNVs bringing about a higher risk for developing sporadic cases of autism. We proposed to test 1) the association of MTHFR 677T allele with rate of ASD related de novo CNVs; 2) the association of MTHFR 677T allele with increased level of global hypomethylation; and 3) the association of level of global hypomethylation with increased rate of ASD related de novo CNVs.					
15. SUBJECT TERMS- nothing listed					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 11	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)

Table of Contents

1. Introduction.....	4
2. Keywords.....	4
3. Overall Project Summary.....	4
4. Key Research Accomplishments.....	4
5. Conclusion.....	5
6. Publications, Abstracts, and Presentations.....	5
7. Inventions, Patents and Licenses.....	5
8. Reportable Outcomes.....	5
9. Other Achievements.....	5
10. References.....	5
11. Appendices.....	attached

INTRODUCTION:

Autism Spectrum Disorder (ASD) are a group of neurodevelopmental disorders that are caused by a range of factors, including: genetic, epigenetic and environmental, with a genetic/epigenetic model proposed (Jiang et al., 2004). While a main focus of autism research remains on the genetic causes, more and more attention was drawn to the role epigenetic factors play, as it has been shown to play a role in idiopathic autism. With our previous published study revealed significantly association of C677T polymorphism in MTHFR gene with idiopathic autism in Simplex (SPX) autism families (Liu et al., 2011); and the proven facts that *de novo* CNVs rates are consistently high in SPX ASD (5.8%-10.2%) versus familial ASD (2-3%), we hypothesize that low-activity MTHFR 677T allele leads to increase global DNA hypomethylation and consequently results in increased generation of *de novo* CNVs bringing about a higher risk for developing sporadic cases of autism. We proposed to test 1) the association of MTHFR 677T allele with rate of ASD related *de novo* CNVs; 2) the association of MTHFR 677T allele with increased level of global hypomethylation; and 3) the association of level of global hypomethylation with increased rate of ASD related *de novo* CNVs.

KEYWORDS

Autism, Sporadic Cases, MTHFR, Hypomethylation, Differentially Methylated Regions (DMR), Copy Number Variation (CNV)

OVERALL PROJECT SUMMARY

This pilot project started from September of 2013. During the over two years of project execution, we have achieved the aims set in the original proposal. Although no publication has yet come out from the study during till today, the significant findings from the study allow us to prepare a significant manuscript that we are aim at submitting to a top journal during this March; and a follow-up study which was carried out with PI's other funding source will very likely yield another manuscript during the middle of this year.

KEY RESEARCH ACCOMPLISHMENTS:

- Completed microarray data analysis on 510 SPX families; identified 99 individuals carrying pathogenic CNVs; confirmed 33 of these CNVs be *de novo*.
- 510 SPX families (both parents and affected individuals) were genotyped on MTHFR Functional Polymorphism C677T using TaqMan Assay.
- Both experimental and bioinformatics pipelines were established for global Methylation profiling using MBD-Seq strategy with Ion Torrent Proton
- 31 ASD cases with *de novo* pathogenic CNVs and 31 ASD cases without *de novo* pathogenic CNVs were MBD-Sequenced. Among all 62 ASD cases, at MTHFR C677T locus, 35, 23 and 4 carry CC, CT and TT genotype respectively.
- The following data analyses were performed
 1. **The MBD-Seq data QA/QC:** Data quality was measured in a number of ways to ensure that the data was of good quality prior to analysis. The first was a measure of saturation, with a cutoff minimum score of 0.5, to determine if the data creates reproducible coverage of the reference genome. The second measure was enrichment, with a minimum cutoff score of 1.7, which measures how well the methylation capture method worked and therefore how enriched the sample is for methylated DNA. The third measure was 5 times or greater coverage, with a minimum cut-off of 5%, meaning that at least 5% of the sequence needed to have at least 5 times sequencing coverage. The fourth measure was the number of total sequencing reads for a sample, with a minimum of 20 million reads. The fifth measure was the number of unique reads for a sample, with a minimum of 15

million reads as a cutoff, where a unique read was defined as a read that does not share a stop or start point with any other read. The number of reads, both total and unique, were plotted to visualize all samples individually to ensure they met this criteria. **Overall, all 62 samples passed all five quality measurements and, thus, were included in the analysis.**

2. ***Global Methylation Index (GMI):*** The global methylation index (GMI) was calculated across the genome for each sample. The average GMI for each intended comparison group, such as group carrying CNV and group not carrying CNV; group carrying genotype C/C and group carrying C/T and T/T genotype on MTHFR C677T polymorphism, was calculated, were compared between groups using t-tests. While no significant difference was found in global methylation level between CNV+ and CNV- groups, we did find the low activity T allele ASD carrier for MTHFR C677T variant has a significantly lower global methylation level than the CC homozygous group.
3. ***Detection of Differentially Methylated Regions (DMRs) between C/C and C/T+T/T groups:*** The methylation level within a 200bp window was calculated and normalized to reads per kilobase per million (RPKM). Following a comparison of RPKM values in each 200 bp window between C/C and C/T+T/T groups at p=0.05, 238 differentially methylated regions (DMRs) were identified. 140 DMRs were hypermethylated in the C/C group relative to the C/T+T/T group. 98 DMRs were hypermethylated in the C/T+T/T group relative to the C/C group. The RPKM values for the DMRs that were found to be hypermethylated in the C/T+T/T group were, generally, lower overall when compared to those identified as hypermethylated in the C/C group. ***It is of great interest that the identified DMRs were not evenly distributed across chromosomes.***
4. ***Most Significantly: The differentially methylated regions (DMRs) were found to be highly biased towards autism related genes and CpG islands.*** This may imply a major mechanism for etiology of sporadic cases of autism: autism causing environmental factors serves as modulator to regulate a MTHFR -mediated epigenomics to regulate specific autism related gene.

CONCLUSION:

The proposed research aims for the funded pilot projects were successfully realized with a significant finding for a potential major mechanism for etiology of sporadic cases of autism.

PUBLICATIONS, ABSTRACTS AND PRESENTATIONS: N/A

INVENTIONS, PATENTS AND LICENSES: N/A

REPORTABLE OUTCOMES:

The significant findings from the study allow us to prepare a significant manuscript that we are aim at submitting to a top journal during this March; and a follow-up study which was carried out with PI's other funding source will very likely yield another manuscript during the middle of this year.

OTHER ACHIEVEMENTS: N/A

REFERENCES:

Jiang YH, Sahoo T, Michaelis RC, Bercovich D, Bressler J, Kashork CD, et al. A mixed epigenetic/genetic model for oligogenic inheritance of autism with a limited role for UBE3A. *Am J Med Genet.* 2004;131A(1):1–10.

Liu X, et al . Population- and family-based studies associate the MTHFR gene with idiopathic autism in simplex families. *J Autism Dev Disord.* 2011 Jul;41(7):938-44.

APPENDENCIES

Findings stated in the progress report and that were to included in the manuscript

MBD-Seq Data Quality

Data quality was measured in a number of ways to ensure that the data was of good quality prior to analysis. The results of these measurements for each sample are summarized in Table 1. The first was a measure of saturation, with a cutoff minimum score of 0.5, to determine if the data creates reproducible coverage of the reference genome (Lienhard *et al.*, 2014). The second measure was enrichment, with a minimum cutoff score of 1.7, which measures how well the methylation capture method worked and therefore how enriched the sample is for methylated DNA. The third measure was 5 times or greater coverage, with a minimum cutoff of 5%, meaning that at least 5% of the sequence needed to have at least 5 times sequencing coverage. Figure 1 is an example of data quality output, while Figure 2 demonstrates the distributions of the quality metrics for each sample group.

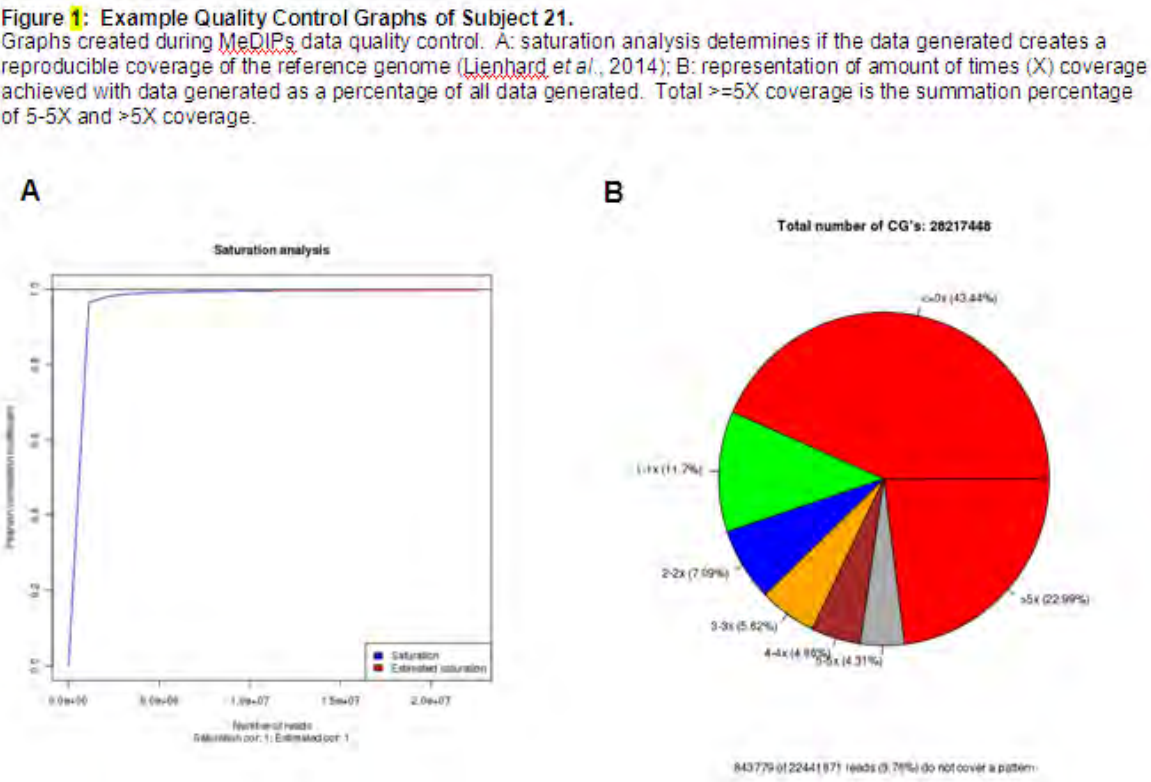


Figure 1: Example Quality Control Graphs of Subject 21.

Graphs created during MeDIPs data quality control. A: saturation analysis determines if the data generated creates a reproducible coverage of the reference genome (Lienhard *et al.*, 2014); B: representation of amount of times (X) coverage achieved with data generated as a percentage of all data generated. Total $\geq 5X$ coverage is the summation percentage of 5-5X and $>5X$ coverage.

Figure 2: Data Quality Metrics Grouped by Genotype

Boxplots of quality control measures of the data grouped by rs1801133 (C677T) genotype groups. Each sample data needed to pass all three thresholds to be considered for analysis. A: percentage of at least 5 times coverage (minimum = 5%); B: relative methylation enrichment score (minimum = 1.7); C: saturation score (minimum = 0.5).

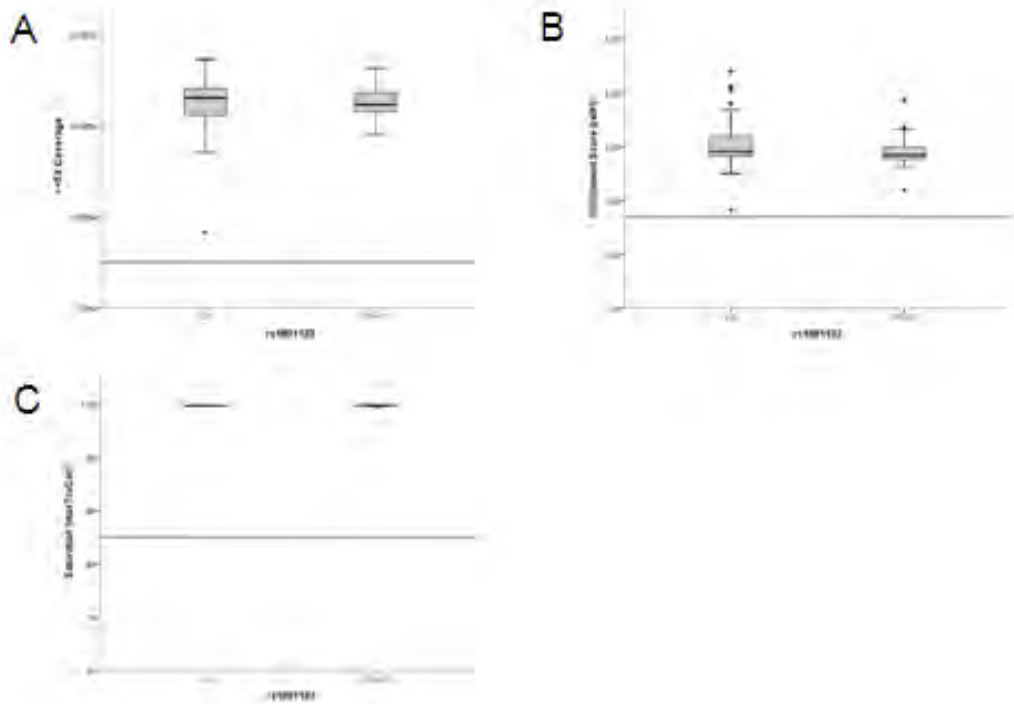


Figure 2: Data Quality Metrics Grouped by Genotype

Boxplots of quality control measures of the data grouped by rs1801133 (C677T) genotype groups. Each sample data needed to pass all three thresholds to be considered for analysis. A: percentage of at least 5 times coverage (minimum = 5%); B: relative methylation enrichment score (minimum = 1.7); C: saturation score (minimum = 0.5).

The fourth measure was the number of total sequencing reads for a sample, with a minimum of 20 million reads. The fifth measure was the number of unique reads for a sample, with a minimum of 15 million reads as a cutoff, where a unique read was defined as a read that does not share a stop or start point with any other read. The number of reads, both total and unique, were plotted to visualize all samples individually to ensure they met this criteria; this can be seen in Figure 3.

Overall, all 62 samples passed all five quality measurements and, thus, were included in the analysis.

Figure 3: Total Reads and Unique Reads

At least 20 mil reads were needed, with at least 15 mil of those reads being classified as unique, to pass quality control.



Figure 3: Total Reads and Unique Reads

At least 20 mil reads were needed, with at least 15 mil of those reads being classified as unique, to pass quality control.

Global Methylation Index (GMI)

The global methylation index (GMI) was calculated across the genome for each sample. The average GMI for each genotype group (C/C, C/T and T/T) was calculated, were compared using t-tests. Following this preliminary analysis, it was decided that the C/T and T/T groups could be combined into one group for further analysis because there was no significant difference in mean GMI between the C/T and T/T (data not shown).

The mean GMI for C/C genotype group was 47.22, with a standard deviations of 18.41 and n=35. The mean GMI for C/T+T/T genotype group was 40.22, with a standard deviation of 9.37 and n=27. Comparing the mean GMI between C/C genotype group and the C/T+T/T genotype group using a t-test showed that there was a marginally statistically significant difference between the groups (p=0.0569; Figure 4).

Figure 4: Mean Global Methylation Index (GMI) for Whole Genome Methylation Comparison Between MTHFR C677T Genotype Groups
 Samples were grouped into two groups: C/C genotype and genotypes with a T present (C/T and T/T). A two-tailed t-test, with independent samples and unequal variances was run using IBM SPSS Statistics v22 to compare the mean global methylation index (GMI), where n=35 for C/C and n=27 for C/T+T/T. *: significantly different, p=0.0569.

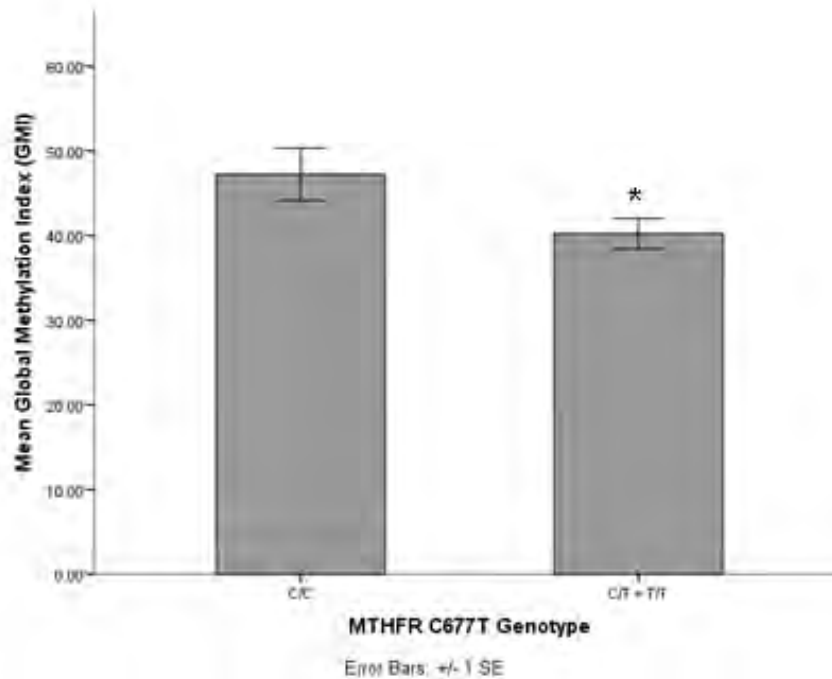


Figure 4: Mean Global Methylation Index (GMI) for Whole Genome Methylation Comparison Between MTHFR C677T Genotype Groups

Samples were grouped into two groups: C/C genotype and genotypes with a T present (C/T and T/T). A two-tailed t-test, with independent samples and unequal variances was run using IBM SPSS Statistics v22 to compare the mean global methylation index (GMI), where n=35 for C/C and n=27 for C/T+T/T.

Analysis of Differentially Methylated Regions (DMRs)

The methylation level within a 200bp window was calculated and normalized to reads per kilobase per million (RPKM). Following a comparison of RPKM values in each 200 bp window between C/C and C/T+T/T groups at p=0.05, 238 differentially methylated regions (DMRs) were identified. These identified DMRs are summarized in a table not presented here. 140 DMRs were hypermethylated in the C/C group relative to the C/T+T/T group. 98 DMRs were hypermethylated in the C/T+T/T group relative to the C/C group. The RPKM values for the DMRs that were found to be hypermethylated in the C/T+T/T group were, generally, lower overall when compared to those identified as hypermethylated in the C/C group (see Figure 5).

Figure 5: Reads per Kilobase per Million (RPKM) Mean Comparison Between Genotype Groups Across Differentially Methylated Regions (DMRs)
 Comparison of mean RPKM values across DMRs when C/T + T/T genotype group was hypermethylated (blue dots) and when C/C genotype group mean was hypermethylated (orange dots).

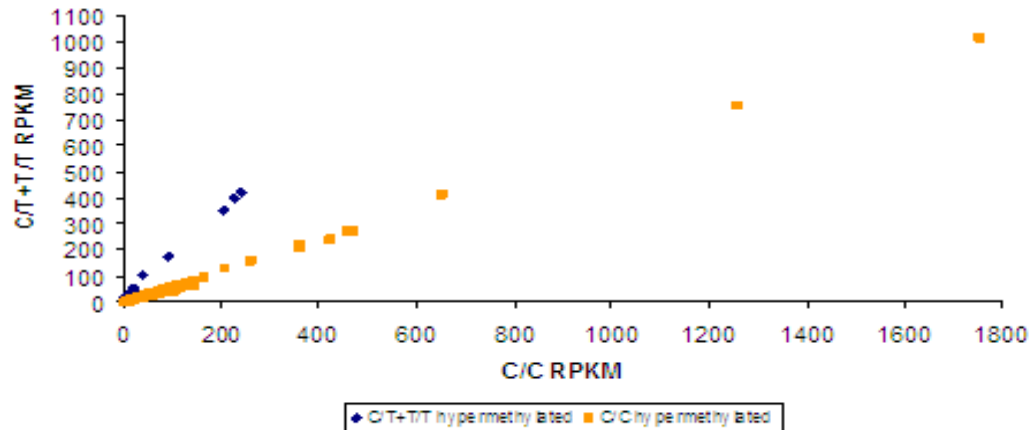


Figure 5: Reads per Kilobase per Million (RPKM) Mean Comparison Between Genotype Groups Across Differentially Methylated Regions (DMRs)

Comparison of mean RPKM values across DMRs when C/T + T/T genotype group was hypermethylated (blue dots) and when C/C genotype group mean was hypermethylated (orange dots).

We found that the identified DMRs were not evenly distributed across chromosomes. This phenomenon was quantified by comparing the actual number of DMRs per chromosome identified to a calculated number of DMRs expected for each chromosome based on chromosome size. It was found that 16 of the 24 chromosomes had a significantly different actual number of DMRs compared to the expected number of DMRs (see Figure 7). Six of the chromosomes were found to be significantly enriched for DMRs while 10 of the chromosomes were found to be reduced in DMRs.

Figure 7: Genomic Distribution of Differentially Methylated Regions (DMRs)
The actual number of DMRs for each chromosome at $p=0.05$ versus the number of DMRs expected per chromosome if the DMRs were distributed evenly based on chromosome size (hg19/GRCh37 Feb 2009; <http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/data/>). There was no data available for mitochondrial DNA.
*: significantly different from expected, $p<0.05$

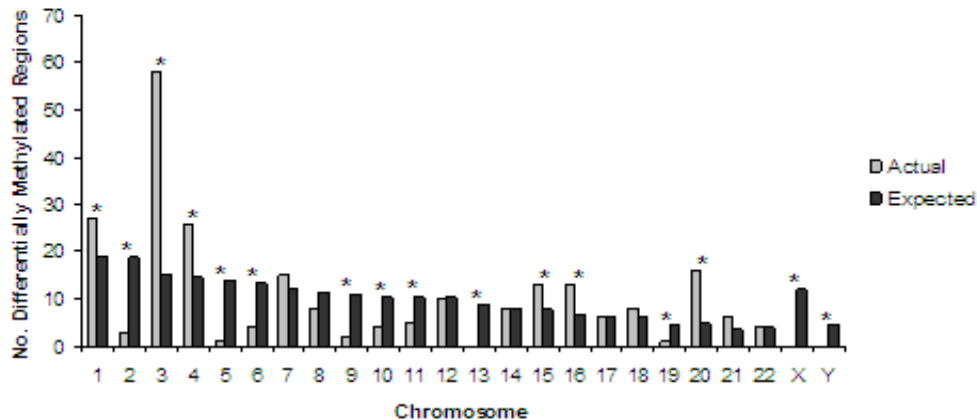


Figure 7: Genomic Distribution of Differentially Methylated Regions (DMRs)

The actual number of DMRs for each chromosome at $p=0.05$ versus the number of DMRs expected per chromosome if the DMRs were distributed evenly based on chromosome size (hg19/GRCh37 Feb 2009; <http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/data/>). There was no data available for mitochondrial DNA.

*: significantly different from expected, $p<0.05$

Sequenom Validation

To validate the findings by MBD-Seq, an alternate method – Sequenom MassARRAY – was chosen to analyze a selection of the DMRs identified. Of the 59 DMRs identified in intragenic regions, 8 of these were selected for validation. Six of the DMRs were hypermethylated in the C/C group and two were hypermethylated in the C/T+T/T group. The data is under analysis.